

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

***MAKING OF* DE NOTRE REPORTAGE DE DONNÉES**

**TRAVAIL FINAL PRÉSENTÉ À L'UNIVERSITÉ DU QUÉBEC À MONTRÉAL
COMME EXIGENCE DU COURS EDM5240 À JEAN-HUGUES ROY**

PAR OLIVIER FAUCHER FAUO19039709

ÉCOLE DES MÉDIAS

AVRIL 2019

Choix du sujet

Après quelques séances de *brainstorming*, Nicolas, Stéphanie et moi avions de la difficulté à trouver un sujet qui allait nous permettre de récolter des données à analyser. Dans la recherche d'un sujet, d'habitude, on doit évaluer l'intérêt public et les intervenants potentiels. Pour un reportage de données, la contrainte de l'accès à des données et leur pertinence était une nouveauté qui représentait un défi. D'abord, j'ai proposé à Nicolas de faire une analyse des commentaires sous des publications Facebook de médias pour comparer leurs lectorats. Cependant, notre professeur Jean-Hugues Roy nous a ramené sur terre quant à l'utilisation des données sur la plateforme Facebook; cette dernière est très stricte et limite fortement l'accès aux données, même si des publications sont « publiques ».

Nous avons ensuite fait connaissance avec les étudiants de Polytechnique. Dans une conversation Slack, suivie d'une rencontre dans un café, nous avons déterminé qu'il était possible de recueillir les données publiques des professeurs de toutes les universités québécoises pour ensuite les analyser. Nous allions observer si certains domaines étaient davantage dominés par un sexe ou l'autre. Ensuite, nous allions utiliser un outil proposé par les étudiants de la Polytechnique pour déterminer l'origine ethnique du ou de la professeur(e). Finalement, nous allions comparer le nombre de professeurs par expertise dans chaque université, afin de nous donner une idée des domaines les plus importants pour les universités. La banque de professeurs allait aussi pouvoir servir lors de la recherche d'intervenants pour les journalistes.

Collecte des données

Nous avons ouvert un document Google dans lequel nous avons mis la liste de toutes les universités avec leur site respectif. Nous avons séparé les tâches: un étudiant en journalisme par université. La tâche s'annonçait colossale, puisque nous avions plus de 20 sites différents à moissonner. Il fallait adapter notre script à chaque site d'université puisque ces sites sont créés et gérés par l'administration; ils n'ont ainsi à peu près rien en commun dans leur structure et leur design. Sur un site, il fallait trouver une liste de professeurs ou un chemin logique qui se rendait vers des listes de professeur par département, par exemple, afin de construire notre script. Notre script irait ensuite sur la page de profil de chaque professeur afin d'y recueillir certaines informations pertinentes, comme les expertises et le genre, par exemple. Le genre pouvait facilement être déterminé grâce à la présence ou non

de la lettre « e » à la fin du titre d'un ou une professeur(e). Le premier site que j'ai tenté de moissonner était celui de l'Université McGill. J'ai cependant frappé un mur; le site de cette université était un véritable fouillis. Sites externes, pas de liste de professeurs, structure illogique: bref, après quelques vérifications avec les étudiants de Polytechnique, nous avons conclu qu'il était impossible de moissonner ce site. Ensuite, je me suis attaqué au site de l'Université de Montréal. Tout fonctionnait alors bien, car une [liste de professeurs](#) existait. Cependant, puisqu'il s'agissait de mon premier script de moissonnage à vie, j'ai mis beaucoup de temps à le compléter. À la toute fin, j'ai malheureusement manqué de temps pour compléter le script de l'Université de Montréal uniquement en raison d'un blocage qui m'empêchait d'amasser les courriels des professeurs, une colonne qui nous a pourtant été inutile par la suite. J'ai aussi été responsable de recueillir les données des professeurs du [HEC](#) et de l'école [Polytechnique](#). Ces deux scripts ont relativement été faciles à écrire. Il y a eu un léger problème avec les expertises des professeurs de la Polytechnique, mais les étudiants de Polytechnique ont pu trouver la solution.

Outils technologiques et sources

Nous avons eu recours à la méthode de moissonnage sur le web avec la bibliothèque Python BeautifulSoup. En complément des enseignements de Jean-Hugues Roy, j'ai eu recours à [la chaîne YouTube de cet ancien employé de Google](#) qui donne des cours d'introduction à Python. Également, le site [w3schools.com](#), une référence en matière d'instructions en codage, m'a beaucoup aidé. Je l'ai utilisé si souvent que je ne me rappelle pas de tous les tutoriels sur Python que j'ai consultés. Les vidéos de Jean-Hugues qui nous apprennent comment moissonner un site web avec BeautifulSoup m'ont été très utiles, mais j'ai aussi eu recours à [cette vidéo explicative d'une demi-heure](#) pour améliorer ma compréhension. Les tutoriels YouTube m'ont beaucoup appris sur certains logiciels que j'utilise à des fins récréatives. C'est pourquoi j'ai le réflexe d'aller fouiller sur cette plateforme lorsque je veux apprendre quelque chose.

Conclusion

La contrainte de temps et certains problèmes techniques nous ont empêchés d'aller chercher tout ce qu'on prévoyait recueillir au lancement du projet. Des universités ont dû être laissées de côté et certaines informations, dont les coordonnées de professeurs, nous ont finalement été inutiles. Notre observation finale quant à ce projet est que nous avons

surestimé nos capacités et notre temps à y consacrer en nous fixant trop d'objectifs à la fois. Toutefois, nous avons tout de même réussi à produire des analyses intéressantes comme la visualisation sur la diversité ethnique selon le domaine, notamment grâce à l'aide de nos collègues de Polytechnique.